SES102

Machines that reflect us

Building generative AI responsibly

Tanvi Singhal

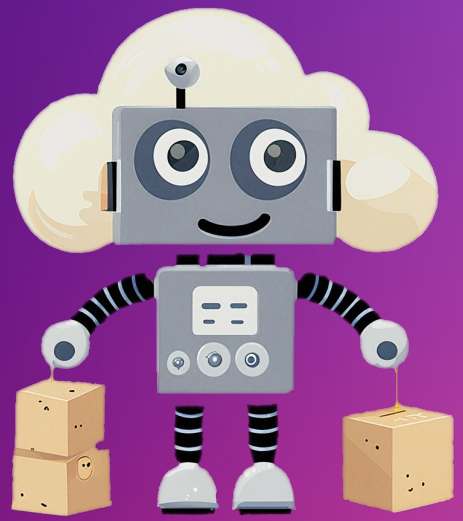
Data Scientist

Aamna Najmi

Senior Data Scientist

Daniela Dorneanu

Manager Data and AI/ML



Why Responsible AI is important?

Example

Prompt: Nurse in a Hospital.

What if you ask differently?

Prompt: Prompt : “Nurse in a Hospital without any indicators of gender identity.”

Why bias matters?

Medical Bias

ESC European Society of Cardiology

European Society of Cardiology › The ESC › ESC

ESC Press Office

NIH National Library of Medicine
National Center for Biotechnolo

PubMed®

Advar

› Acad Emerg Med. 2008 May;15(5):414

Gender disparity in an emergency departmen abdominal pain

* Article from Diagnostic and Interver

uardian

The Guardian

Feature | Womens Cardiovascular Health | June 27, 2022 | By Jennifer Jones-McMeans, Ph.D., DVP

Closing the Gap in Healthcare by Addressing Gender Bias

Achieving greater gender equity in healthcare is about making systemic changes on a comprehensive scale

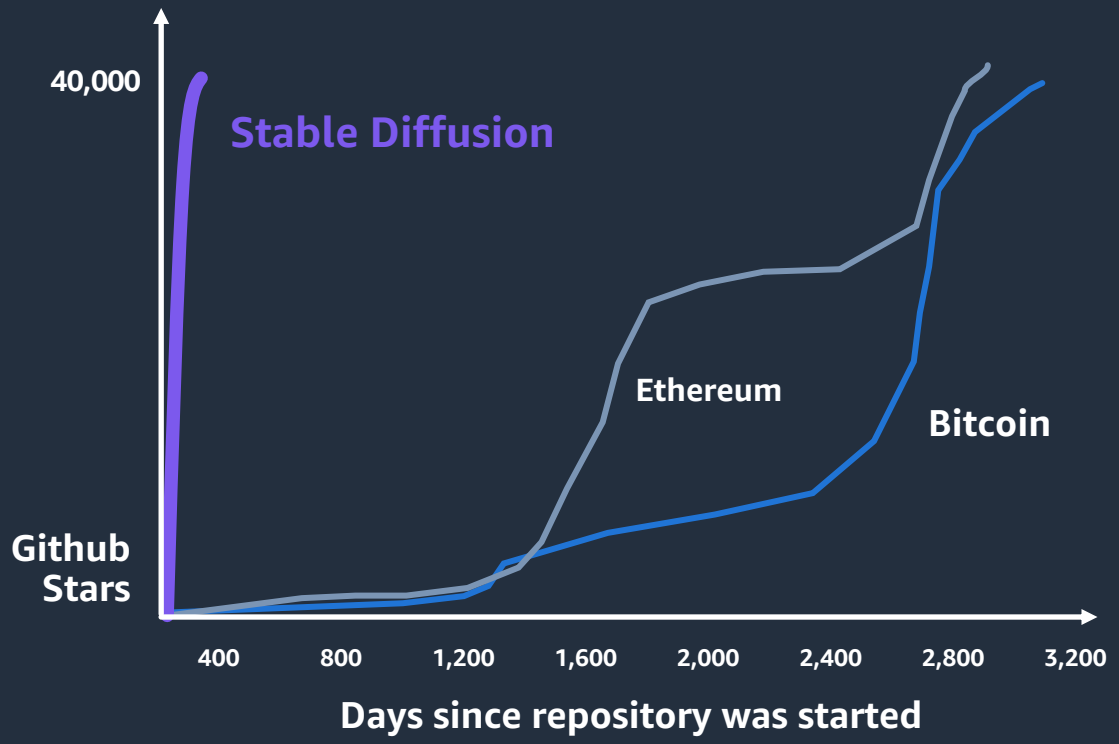
Preclinical Research

NIH National Institutes of Health (NIH)
98K subscribers
Subscribe

Generative AI is the fastest growing trend in AI

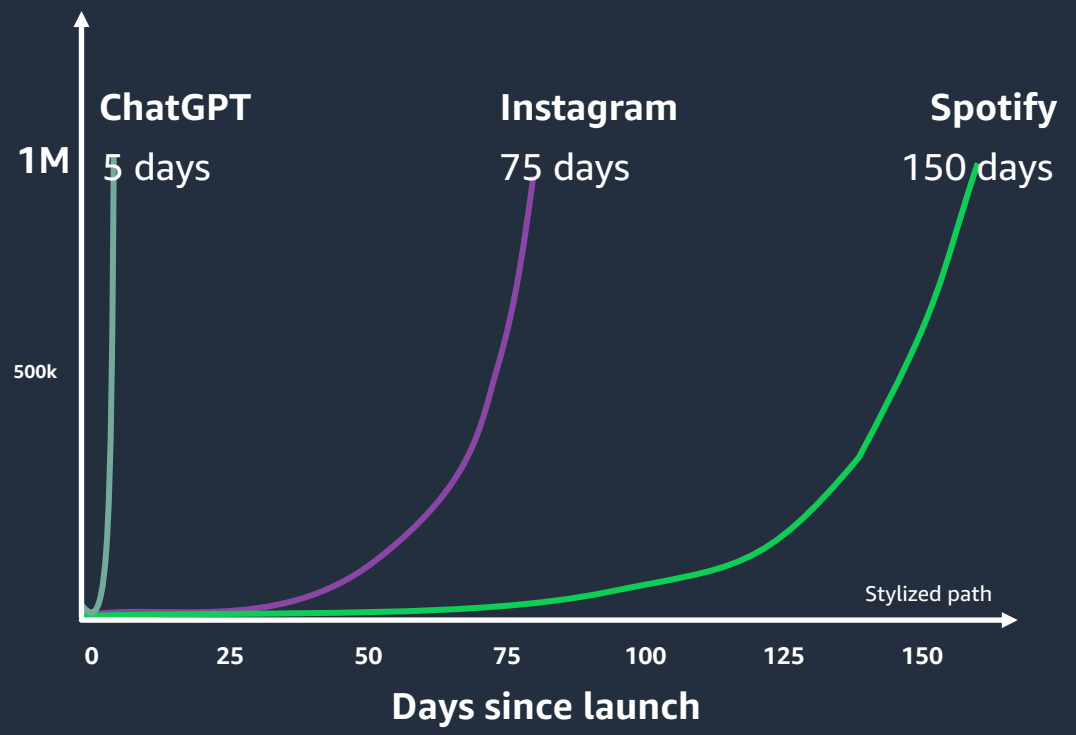
Developer adoption

Stable Diffusion accumulated 40k stars on GitHub in its first 90 days



Consumer adoption

Generative AI chatbot reached the 1 million users mark in 5 days



What is the mental model to build AI responsibly?

Why

1

Know risks and challenges

4 Risks

2

3

Emerging risks and challenges with generative AI

Hallucinations
assertions or claims that sound plausible but are verifiably incorrect.

Toxicity & safety
generate offensive, disturbing, inappropriate content

Intellectual property

Data privacy

What is the mental model to build AI responsibly?

Why

1

Know risks and challenges

4 Risks

2

Memorize and use the responsible AI tenets

8 Tenets

3

Responsible AI Considerations

Governance

Incorporating best practices into the AI supply chain, including providers and deployers

Safety

Preventing harmful system output and misuse

Privacy & Security

Appropriately obtaining, using and protecting data and models

Fairness

Considering impacts on different groups of stakeholders

Veracity & Robustness

Achieving correct system outputs, even with unexpected or adversarial inputs

Explainability

Understanding and evaluating system outputs

Transparency

Enabling stakeholders to make informed choices about their engagement with an AI system

Controllability

Having mechanisms to monitor and steer AI system behavior

What is the mental model to build AI responsibly?

Why

1

Know risks and challenges

4 Risks

2

Memorize and use the responsible AI tenets

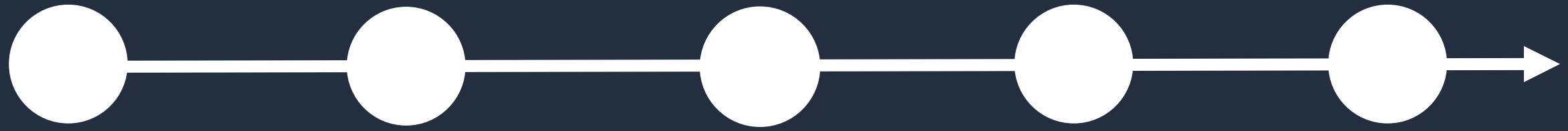
8 Tenets

3

Apply Mitigation techniques

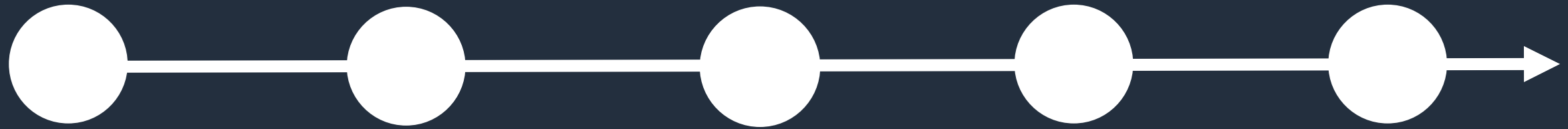
Technical approaches

The ML Lifecycle



Where to consider responsible AI in ML Lifecycle?

Join at
menti.com
and use code
17097561



Business Problem

Pre - Processing

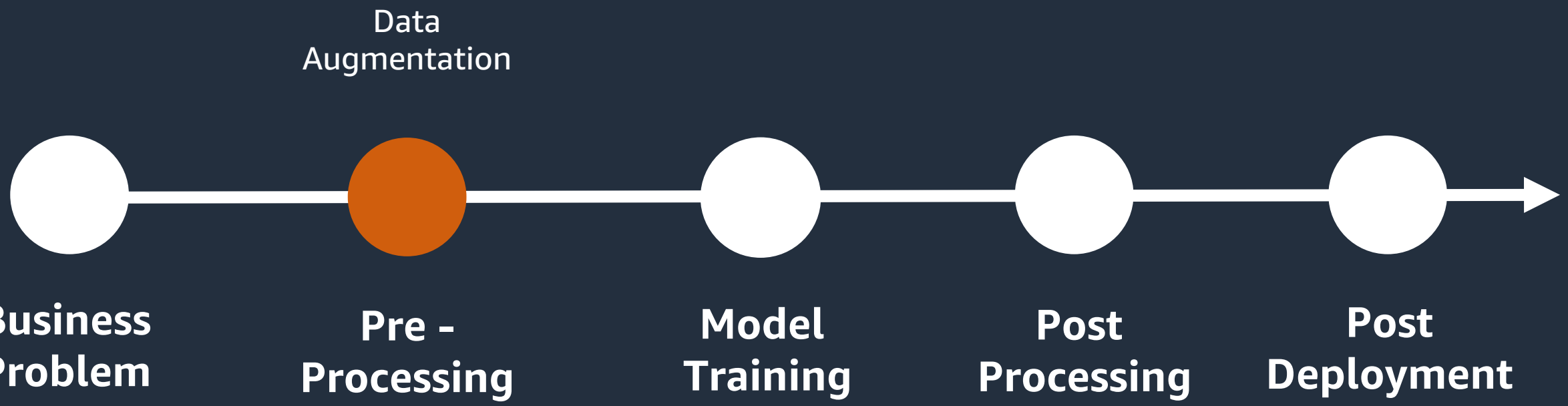
Model Training

Post Processing

Post Deployment

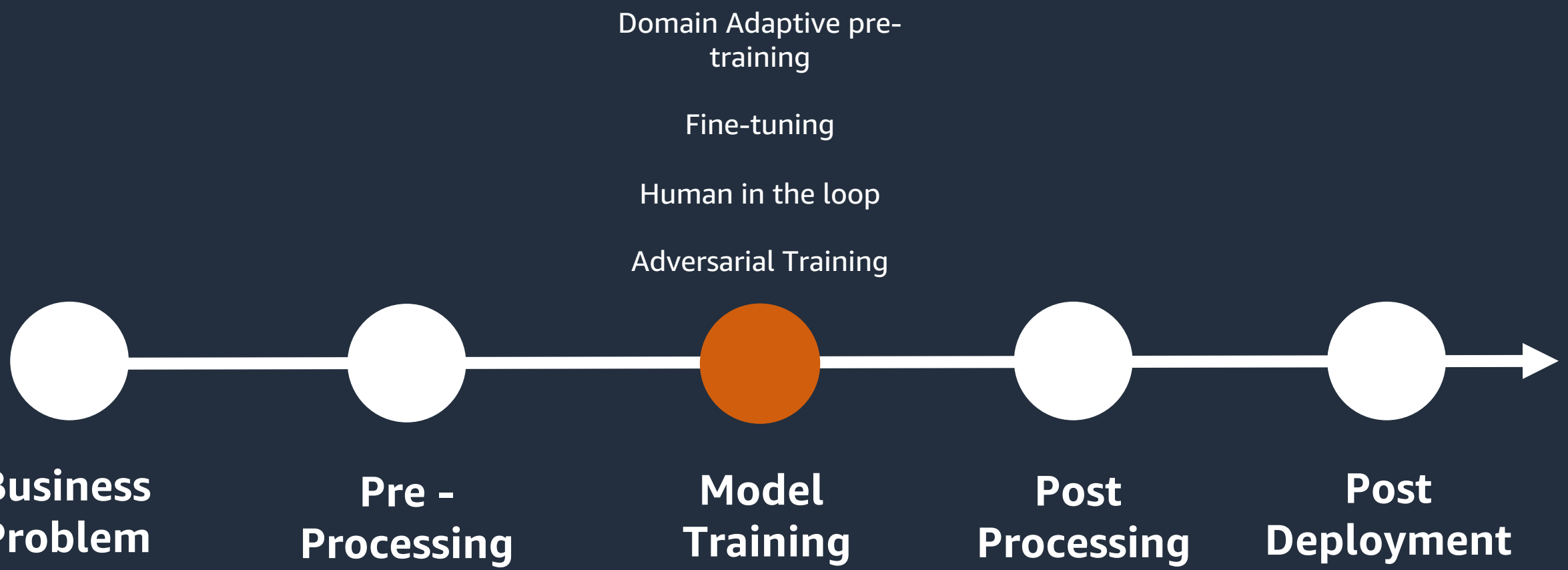
Where to consider responsible AI in ML Lifecycle?

Where to consider responsible AI in ML Lifecycle?



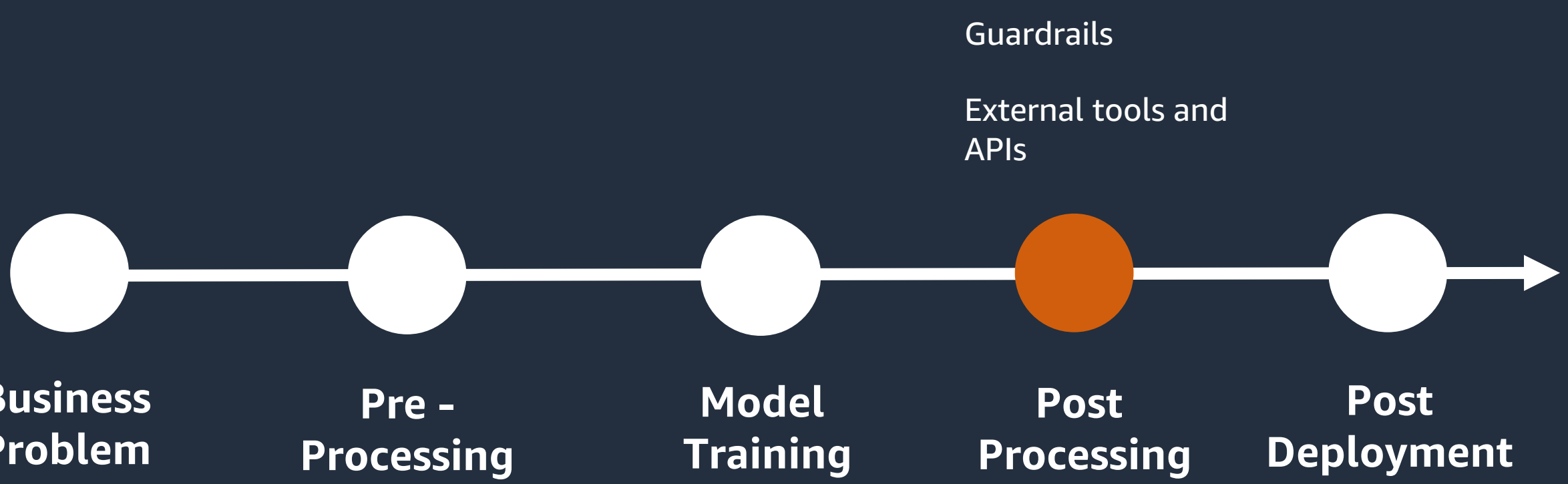
Data that represents all of us

Where to consider responsible AI in ML Lifecycle?

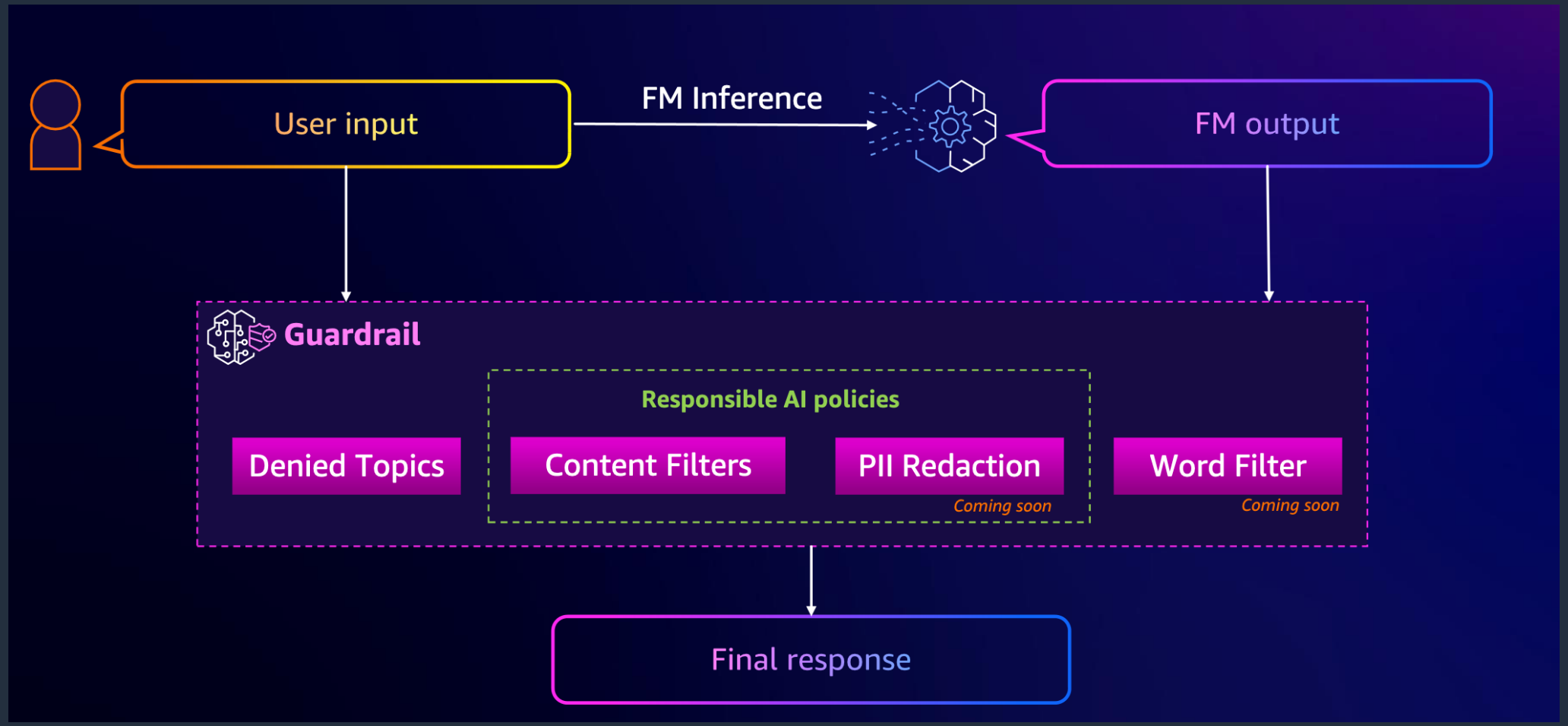


Building ML workloads is like pizza making

Where to consider responsible AI in ML Lifecycle?

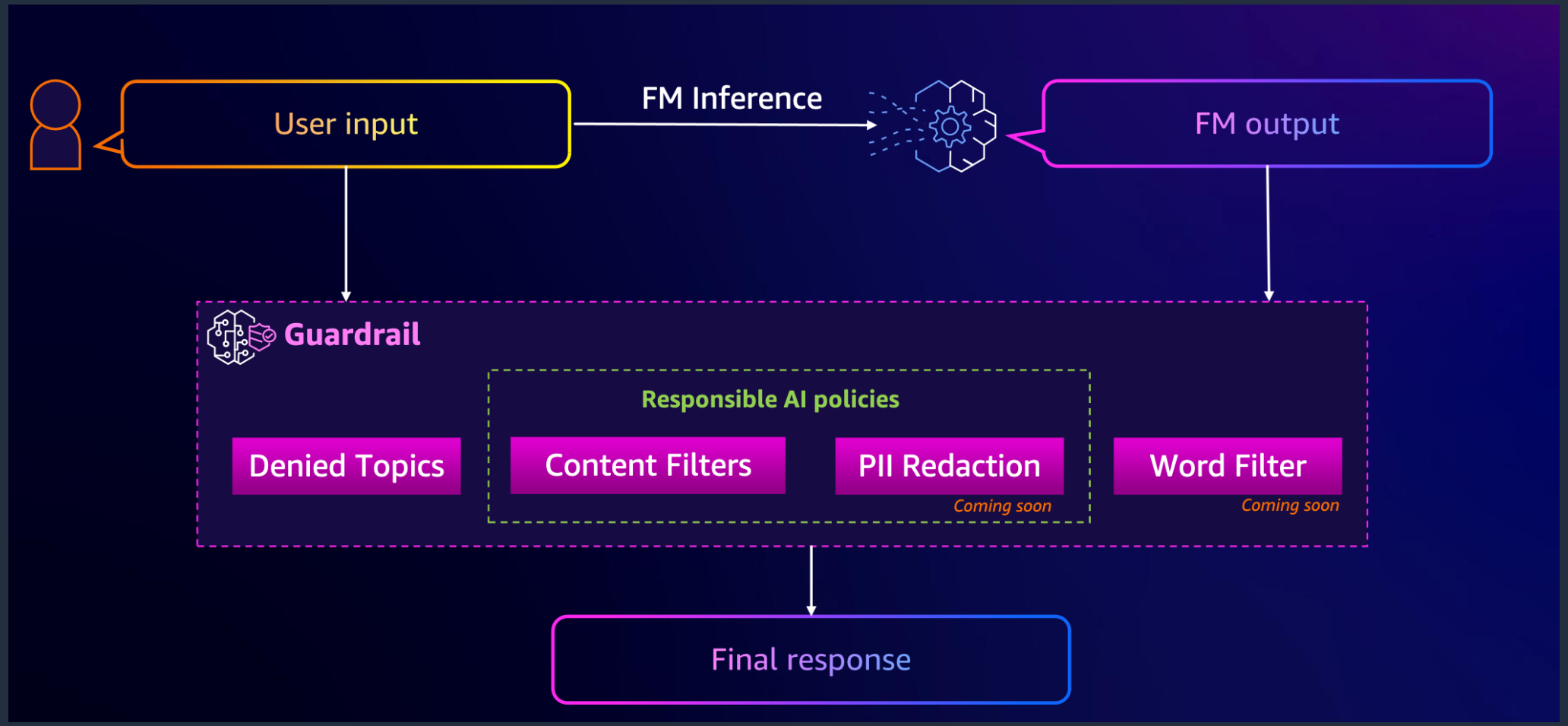


Guardrails in action

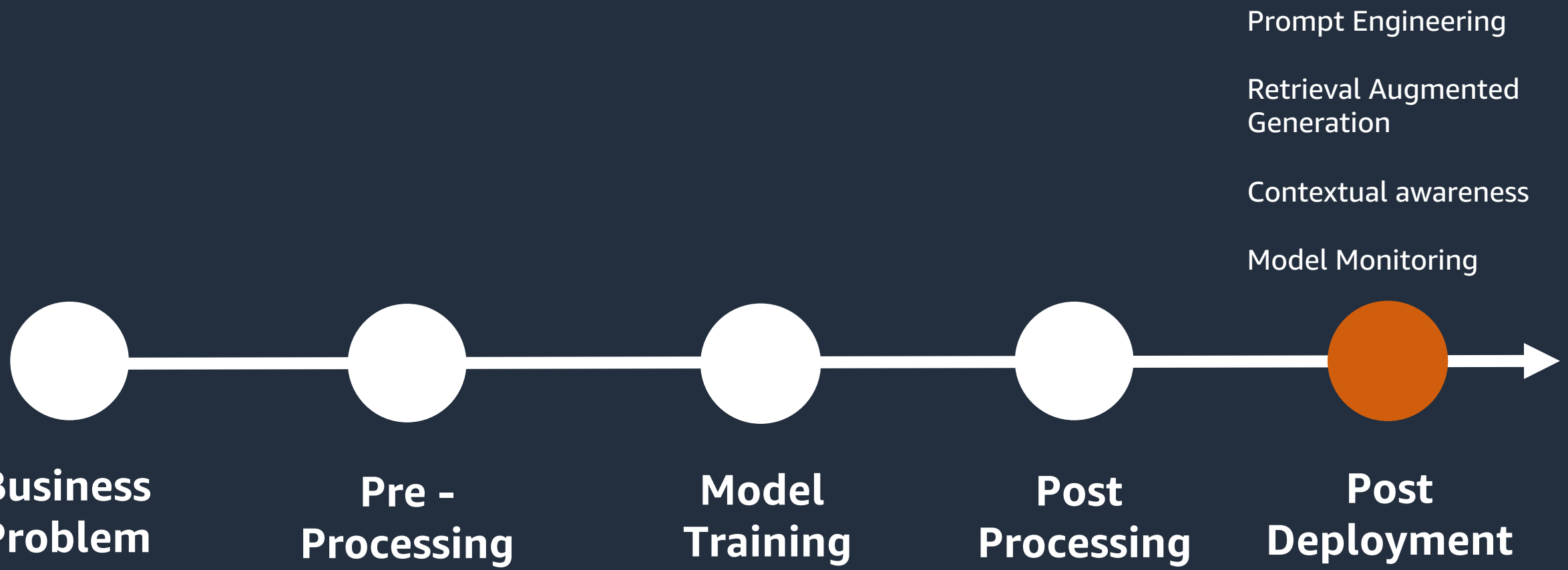


Guardrails in action

Guardrails will ensure you won't ever have pineapple on your pizza



Where to consider responsible AI in ML Lifecycle?



Model Post-Processing with Prompt Engineering

Text-to-Image Disambiguation (TIED) Framework - *Aims to disambiguate the prompts given to the text-to-image generative models by soliciting clarifications from the end user* - Mehrabi et al 2023

Ambiguous
Prompt: An image of a diplomat.




Disambiguous
Prompt: An image of a diplomat. The diplomat X.

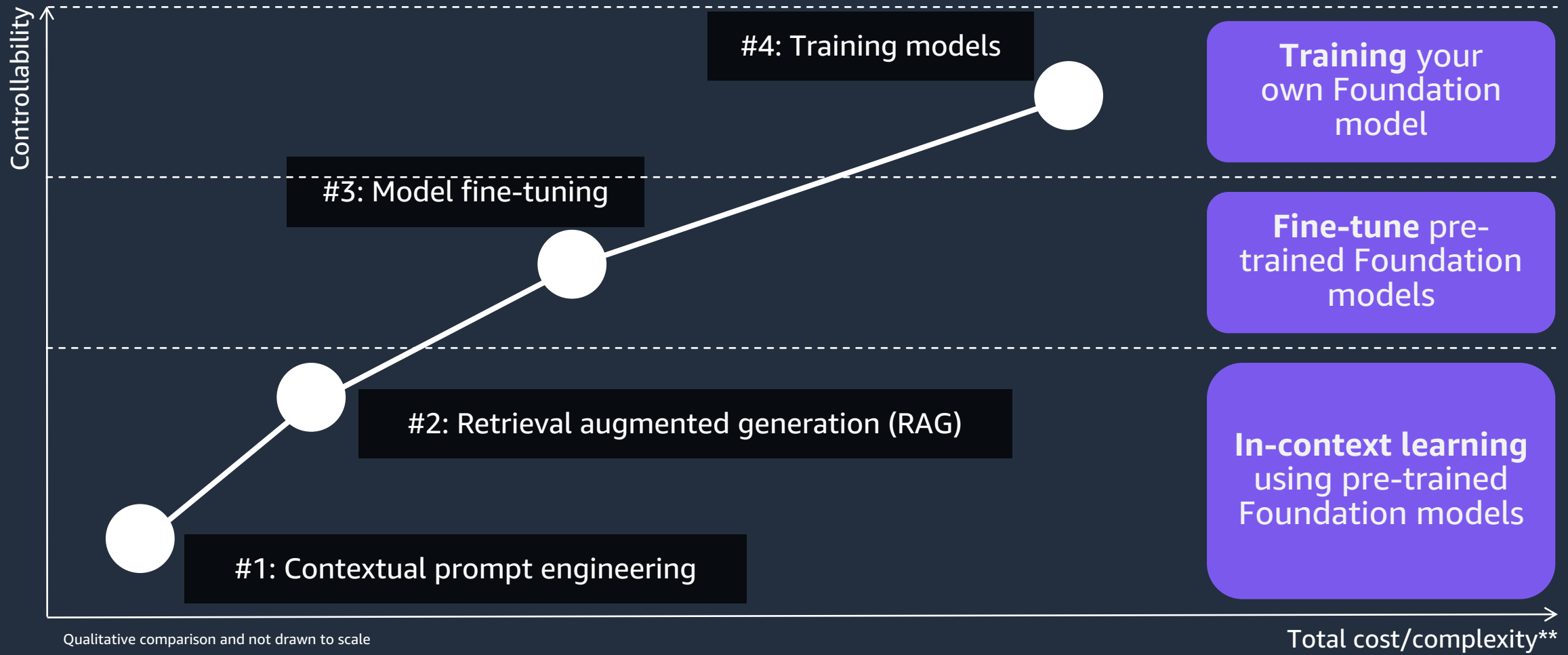
X= is a male

X= is a female

X= is an old man

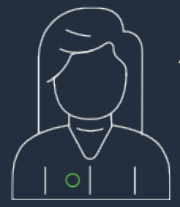
X= is a female with dark skin color.

Generative AI model implementation patterns



*** Total cost/complexity of model preparation, inference, Data management*

Customer Stories



Pet Owner

"Recommend me a product for my pet"

The recommended product is XYZ and it was picked because

Amazon Bedrock

Prompt Engineering with user profile as context

Knowledge Base using the product catalog

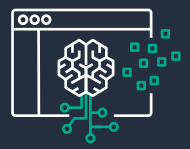
Parameter Tuning

Guardrails

On-Call Sales Agent

What is Mallorca Policy in HDI?

X's Carefree vacation with the Mallorca policy […]



Chatbot powered by LLM on Amazon Bedrock

Prompt Engineering and context awareness

Knowledge Base using the policy documents

Parameter Tuning

What is the mental model to build AI responsibly?

Why

1

Know risks and challenges

4 Risks

2

Memorize and use the responsible AI tenets

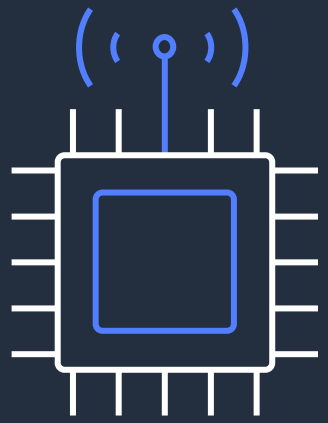
8 Tenets

3

Apply Mitigation techniques

Strategic Approaches

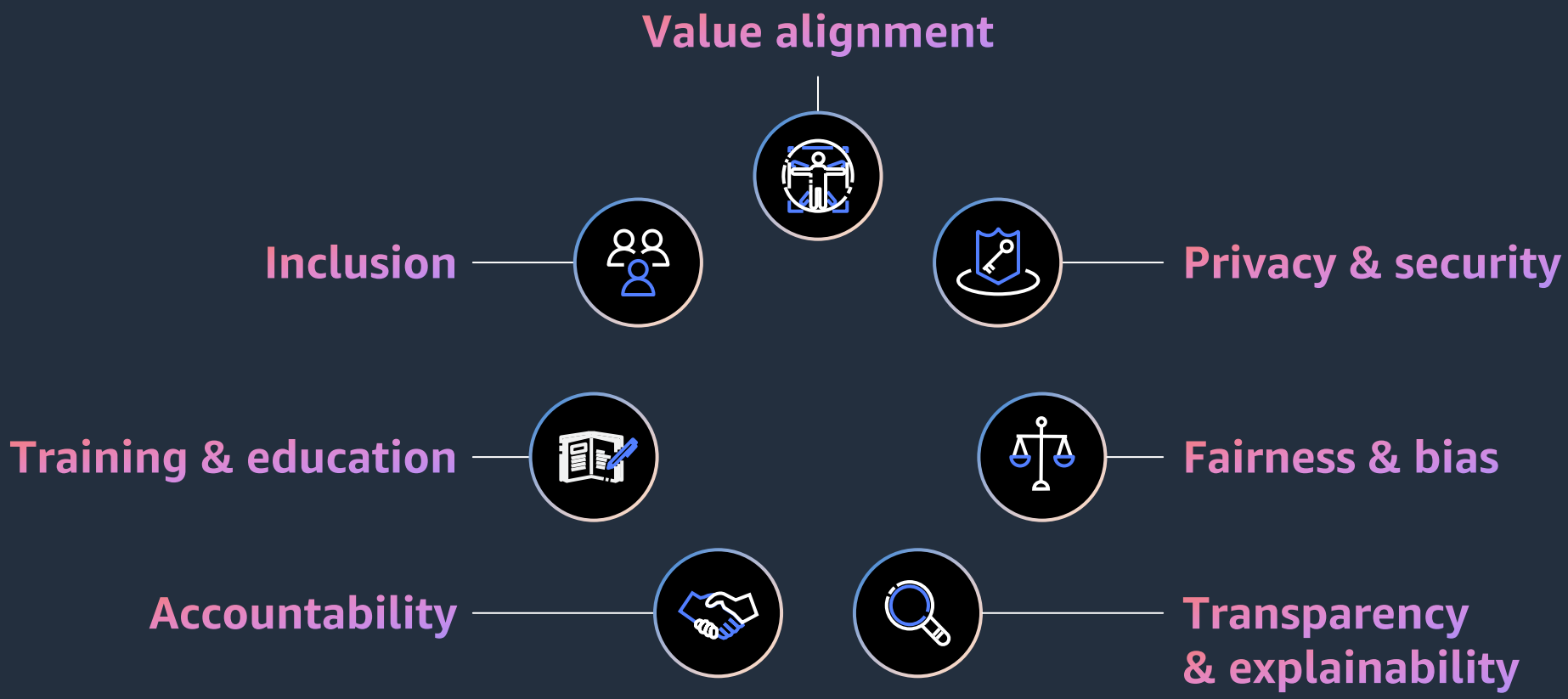
Making the shift



Technology

Organizational

Strategic Approach



Services and features to help build responsibly

Accountability

Amazon Augmented AI (A2I)

Amazon SageMaker
ML Lineage Tracking

Amazon SageMaker
Experiments

Privacy & security

Amazon Macie

AWS Lake Formation

Amazon CodeWhisperer

AWS Glue DataBrew

AWS security/identity services

Bias & fairness

Amazon SageMaker Clarify

Amazon SageMaker
Model Monitor

Amazon SageMaker
Data Wrangler

Amazon SageMaker Autopilot

Amazon Augmented AI (A2I)

Explainability

Amazon SageMaker Clarify

Amazon SageMaker
Model Monitor

Amazon SageMaker
Model Registry

Amazon SageMaker
ML Lineage Tracking

Recap

Why

1

Know risks and challenges

4 Risks

2

Memorize and use the responsible AI tenets

8 Tenets

3

Apply Mitigation techniques

2 approaches

Call to Action

Read about **Responsible AI** from specialists.

Watch free, public course on fairness criteria and bias mitigation in the ML lifecycle.

Take the next steps in your responsible AI journey

Contact us for your complimentary assessment: https://aws.amazon.com/responsibleai

How would you mitigate risks in AI and beyond?

Let’s generate our vocabulary (1-3 words)

or

Join at
menti.com

and use code
17097561

Thank you!

https://pulse.aws/survey/SCDDHVUH